

Network Filesystems and Grid Computing

Matthias Danner

June 18, 2013

- 1 Network Filesystems
 - Introduction
 - Samba
 - Network File System
 - Other distributed filesystems

- 2 Grid Computing
 - Introduction
 - Open Grid Scheduler

1 Network Filesystems

- Introduction
- Samba
- Network File System
- Other distributed filesystems

2 Grid Computing

- Introduction
- Open Grid Scheduler

Use cases of Network Filesystems

- Data sharing between multiple users
- Access to data from different machines
- Usage of redundant data storage
- Opportunity to use diskless workstations

Disadvantages of Network Filesystems

- Network traffic
- Personal data insecure if not proper installed

Samba

- free re-implementation of SMB (Server message block) protocol¹ by Andrew Tridgell
- runs on most Unix and Unix-like machines
- provides Windows' printer and file services
- can serve as Windows Primary Domain Controller

¹also known as CIFS (Common Internet File System) protocol

Samba configuration

Configuration file: `/etc/samba/smb.conf`

- structured in sections

Three special sections: `global`, `home`, `printer`

Customized sections can be added. Common parameters are:

`comment` hold the label to be given to the share

`path` specify the full path of the share directory

`browseable` whether the client can list the content of the share directory

`writable` whether the client can write something to the share directory

`valid user` define who can access the share sources

use `testparm` to check syntax

Samba configuration

Samba provides four security levels:

- `share` corresponds to Windows' access control

- `user` corresponds to Linux' access control

- `server` like user, but uses another server to authenticate

- `domain` authentication is forwarded to Windows Primary Domain Controller

Network File System

Definition

Network File System (NFS) is a distributed file system protocol originally developed by Sun Microsystems in 1984, allowing a user on a client computer to access files over a network in a manner similar to how local storage is accessed.^a

^ahttp://en.wikipedia.org/wiki/Network_File_System

NFS configuration

Configuration file: `/etc/exports`

After changing `/etc/exports` inform server via `exportfs`

Each line indicates a shared volume.

`/etc/exports` Syntax:

`<path> <computer>[<options>] [<computer2>[<options>]]`

`path` full path of shared volume

`computer` IP or domain name (wildcards can be used)

`options` definition of access rights, ...²

²check man `/etc/exports`

Other distributed filesystems

Availability of many other distributed filesystems:

- sshfs
 - uses ssh to connect to remote server
 - mounted folder is easily accessible
- OCFS2
 - combines multiple hosts to serve out a single filesystem
 - enables concurrent access to filesystem
- GlusterFS
 - aggregates multiple storage servers
 - redundant storage

1 Network Filesystems

- Introduction
- Samba
- Network File System
- Other distributed filesystems

2 Grid Computing

- Introduction
- Open Grid Scheduler

Grid Computing

Definition

Grid computing is the federation of computer resources from multiple locations to reach a common goal. [...] Grids are often constructed with general-purpose grid middleware software libraries.

[...] Grids are a form of distributed computing whereby a "super virtual computer" is composed of many networked loosely coupled computers acting together to perform large tasks.^a

^ahttps://en.wikipedia.org/wiki/Grid_computing

Grid Computing

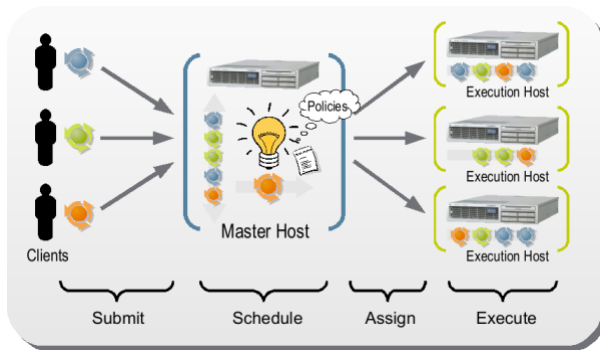
Definition

Grid computing is the federation of **computer resources** from **multiple locations** to reach a common goal. [...] Grids are often constructed with general-purpose grid middleware **software libraries**.

[...] Grids are a form of distributed computing whereby a "super virtual computer" is composed of many **networked loosely coupled** computers acting together to perform large tasks.^a

^ahttps://en.wikipedia.org/wiki/Grid_computing

Grid Computing



Open Grid Scheduler

- originally developed under the name Sun Grid Engine
- in 2001 it became open source
- in 2010 after Oracle purchased Sun the Community started the open source project Open Grid Scheduler
- normally used for high performance computer clusters
- used for the BIM compute cluster

Advantages of the OGS

- manages resources
 - user can specify resources (e.g. memory, cores)
 - job runs only if slot is free and the job has the highest priority
- shared filesystem is mounted
 - Output- and Errorstream are redirected to user's home directory
 - source files and data can be stored in user's home directory

References

Samba:

- http://wiki.samba.org/index.php/Main_Page
- [de] <http://wiki.ubuntuusers.de/Samba>

Network File System:

- <http://nfs.sourceforge.net/nfs-howto/>
- [de] <http://wiki.ubuntuusers.de/NFS>

Open Grid Scheduler:

- <http://gridscheduler.sourceforge.net/howto/GridEngineHowto.html>

Last year's presentation from Christian Mertes

Thanks for your attention!
Questions?